Volume 9 Number 16 1981 Nucleic Acids Research

# Reiterated sequences within the intron of an immediate-early gene of herpes simplex virus type 1

Roger J.Watson, Kenichi Umene and Lynn W.Enquist

Laboratory of Molecular Virology, National Cancer Institute, Bethesda, MD 20205, USA

Received 22 April 1981

## ABSTRACT

We describe the nucleotide sequence of a herpes simplex virus type 1 DNA fragment containing the intron of the immediate-early mRNA-5 (IE mRNA-5) gene. The location of the intron within this fragment was determined by a Berk & Sharp nuclease S1 protection analysis, and by cloning and sequencing cDNA containing sequences overlapping the IE mRNA-5 splice point. We found that the 149 base pair (bp) intron contained four copies of an identical 23 bp GC rich tandem repeat followed by a further reiteration consisting of the first 15 bp only.

## INTRODUCTION

The five predominant immediate-early mRNAs (designated IEmRNA-1 through -5) that accumulate in the cytoplasm of cells infected with herpex simplex virus type 1 (HSV-1) in the presence of cycloheximide, constitute the earliest class of HSV-1 transcripts so far detected (1-3). The locations of the genes encoding the immediate-early mRNAs have been mapped on the virus genome (4) and their directions of transcription have been described (5). Subsequently, we reported (6) that the genes encoding IEmRNA-4 and IEmRNA-5 contain a single intron of approximately 150 base pairs (bp) mapping at equivalent loci within the two inverted repeats ($TR_S$ and $IR_S$) which flank the short (S) component of the virus DNA. Both IE mRNA-4 and IE mRNA-5 contain 260 base (b) 5' leader sequences which map within $TR_S/IR_S$. These leader sequences are spliced to 3' terminal co-transcripts of 1450 b (for IE mRNA-4) and 1540 b (for IE mRNA-5) that contain a small amount of homology with $TR_S/IR_S$, but which are encoded predominantly by opposite ends of the unique S ($U_S$) DNA component (6). In this report we describe the nucleotide sequences of an SmaI fragment containing the IE mRNA-5 gene intron, and of cDNA containing sequences overlapping the IE mRNA-5 splice point. In this way, we have been able to precisely locate and size the intron. We found that the intron consisted largely of a 23 bp reiterated unit.

## MATERIALS AND METHODS

Labeling DNA fragments: Subfragments of the HSV-1 BamHI 4/EcoRI (B4R1) fragment carried in plasmid pKL43 (6) were $^{32}$P-labeled at the 5' terminus as described by Maxam & Gilbert (7). 3'-$^{32}$Plabeling of DNA restriction fragments containing 5' single-strand overhangs were labeled as described by Chalberg & Englund (8). Fragments containing 3' single-strand DNA overhangs were labeled with terminal transferase and $\alpha$-$^{32}$P-cordycepin 5'-triphosphate using a kit obtained from New England Nuclear Inc., essentially by the procedure of Roychoudhury & Wu (9).

Nuclease S1 analysis: Hybrids between 5'-$^{32}$P-labeled DNA fragments and cytoplasmic immediate-early RNA were formed and treated with nuclease S1 as described previously (6). Ethanol-precipitated RNA/DNA hybrids were resuspended in 20 µl 90% formamide/1 x TBE (1 x TBE is 50 mM Tris, 50 mM boric acid, 1 mM EDTA, pH 8.3) containing dyes, and the hybrids were denatured by heating at 90°C for 2 minutes. Appropriate amounts of each sample were applied to a 1.5 mm thick 8% acrylamide-0.27% bisacrylamide gel cast in 8 M urea, 1 x TBE. The gel was run at 30 V/cm for 1 hr, or until the xylene cyanol dye was approximately 2/3 the way down. The gel was then dried and subjected to autoradiography.

DNA sequencing: Sequencing of end-labeled DNA fragments was performed as described by Maxam & Gilbert (7).

cDNA synthesis and cloning: To synthesize cDNA specific for IE mRNA-5, 0.5 µg of the 110 bp BstNI/SmaI fragment (Figure 1) was hybridized to 10 µg of cytoplasmic polyadenylated immediate-early RNA (of which between 0.5 and 1 µg was IE mRNA-5). Conditions for hybridization and RNA extraction have been described before (6). The hybridized primer was extended by reverse transcriptase, then the RNA was digested with alkali (10). The second strand of the cDNA was synthesized using the Klenow fragment of E. coli DNA polymerase (11), and the products were treated with nuclease S1 (12). The cDNAs were tailed with approximately 10 dC residues using terminal transferase (13), and were annealed to pBR322 cut at the PstI site and tailed with dGTP (12,13). The annealed mixture was used to transform LE392 made competent for DNA uptake (14), and tetracycline-resistant colonies were screened for the presence of sequences complementary to the 280 bp BstNI/SmaI and the SmaI C fragments (Figure 1), using the colony hybridization procedure of Grunstein & Hogness (15).

RESULTS:

Location of the acceptor splice site within the SmaI A fragment: The SmaI map of the BamHI 4/EcoRI (B4R1) fragment of HSV-1 DNA carried in plasmid pKL43 (14) is shown in Figure 1. The B4R1 DNA fragment comprises the rightmost sequences of the Eco RI H fragment (in the prototype orientation) and thus contains both $U_S$ and $TR_S/IR_S$ sequences (6). By DNA sequencing, we have located the junction between unique and reiterated DNA in the 431 bp Sma A subfragment (Watson et al., in preparation): the three Bst NI sites shown in the expanded map of the SmaI A fragment (Figure 1) all lie within $U_S$, whereas the single TaqI site lies within $TR_S/IR_S$.

We demonstrated previously, by nuclease S1 protection analyses (6), that the 3' or acceptor splice site of IE mRNA-5 maps within the B4R1 fragment at a distance of 785 b from the BamHI 4 site. This places the 3' intron-exon junction of the IE mRNA-5 gene in the SmaI A fragment of B4R1 (Figure 1). To locate this intron-exon junction more precisely, a further S1 analysis was performed.

The probes used in this S1 analysis were the SmaI A fragment labeled at both 5' termini, and the 110 and 280 bp BstNI/SmaI subfragments of the SmaI

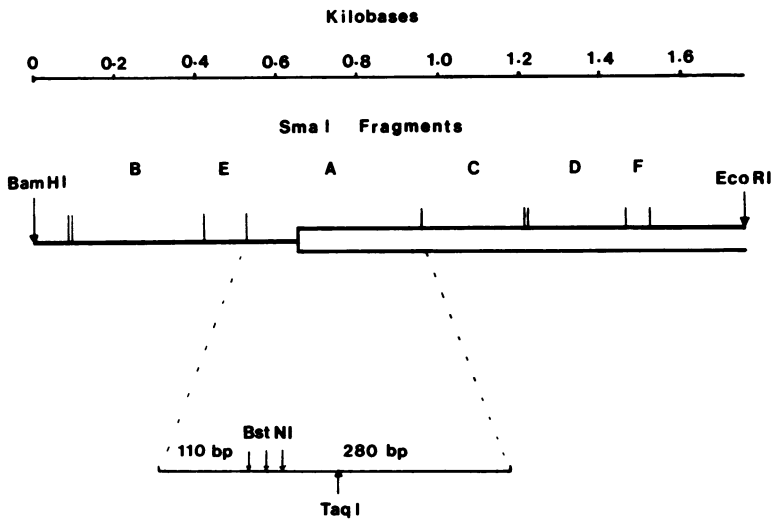


Figure 1. Map of the SmaI cleavage sites within the BamHI 4/EcoRI DNA fragment. The single line represents $U_S$ sequences and the double line $TR_S/IR_S$ sequences. The lower expanded map of the SmaIA fragment shows the BstNI and TaqI cleavage sites and the locations of the 110 and 280 bp SmaI/BstNI subfragments.

A fragment (Figure 1), uniquely labeled at their BstN1 5' termini. The labeled DNA probes were denatured and hybridized to cytoplasmic immediate-early RNA under conditions favouring RNA/DNA duplex formation. Following hybridization, single-stranded nucleic acids were digested with nuclease S1, and the labeled DNA fragments protected by hybrid formation with RNA were analyzed by electrophoresis through a denaturing polyacrylamide gel. Autoradiography of this gel (Figure 2) revealed protection of a 230 b DNA fragment with the Sma1 A probe, and an 80b DNA fragment with the 280 bp Bst N1/Sma1 fragment probe. As expected, the labeled strand of the 110 bp Bst N1/Sma1 fragment did not hybridize to RNA, being of the same polarity as IE mRNA-5. Control experiments, in which these DNA probes were incubated in the absence of immediate-early RNA, indicated that there was no DNA reassociation under the experimental conditions (Figure 2).

The data detailed above indicated that the 3' intron-exon junction of the IE mRNA-5 gene is located 230 b from the left Sma1 site of the Sma1 A fragment, in the orientation indicated in Figure 1, and 80 b from the rightmost BstN1 site. Because the intron size is approximately 150-170 bp (6), it seemed probable that the entire intron is contained within the 431 bp Sma1 A fragment.

Nucleotide sequence of the Sma1 A fragment: The sequences of both strands of the Sma1 A fragment were determined using the chemical degradation method of Maxam & Gilbert (7). Briefly, the strategy involved sequencing 5' and 3' end-labeled Sma1 A fragments secondarily cut with BstN1. The sequence across the BstN1 sites was determined by sequencing from the 5' and 3' end-labeled internal Taq1 site. Isolation of the fragments used in sequencing was facilitated by cloning the Sma1 A fragment in the Pst1 site of plasmid pBR322 by the GC-tailing procedure (13).

An autoradiograph of a gel used to sequence the 280 bp BstN1/Sma1 fragment (3'-$^{32}$P-labeled at the Pst1 site external to the tailed Sma1 site) shown in Figure 3, reveals graphically the presence of a repeated sequence within this fragment. The repeating sequence in this, the template strand, is extremely C rich. The complete sequence of the non-coding strand of the Sma1 A fragment is shown in Figure 4.

The reiterated sequence observed in Figure 3 consists of four full copies of a 23 bp sequence, followed by a partial repeat of the first 15 bp only (residues 63-169, Figure 4). From the nuclease S1 analysis described above, and from data described later in the RESULTS, it was determined that these repeats lie within the IE mRNA-5 gene intron.

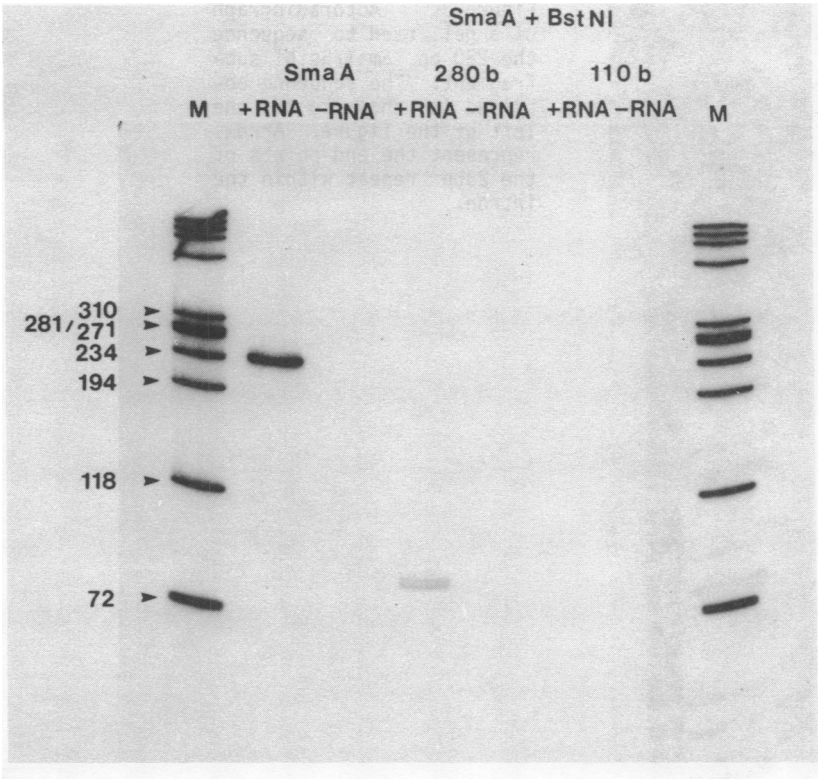


Figure 2. Location of the 3' intron/exon junction using nuclease S1. The SmaIA fragment $^{32}$P-labeled at both 5' termini, and the 280 and 110 bp SmaI/BstNI subfragments uniquely labeled at the 5' BstNI termini, were incubated either with (+) or without (-) cytoplasmic immediate-early RNA and the products were treated with nuclease S1. The sizes of the DNA probes protected by duplex formation with IE mRNA-5 were analyzed by electrophoresis on a urea/polyacrylamide gel using the 5'-$^{32}$P-labeled HaeIII fragments of Øx174 DNA as size markers (M).

Cloning and sequencing cDNA overlapping the IE mRNA-5 splice point: cDNA specific for IE mRNA-5 was synthesized by hybridization of the 110 bp SmaI/BstN1 subfragment of SmaI A (Figure 1), which maps entirely within $U_S$, to cytoplasmic polyadenylated immediate-early RNA and extension of the primer using reverse transcriptase. The complementary strand of cDNA was made using DNA polymerase and the resultant double-stranded cDNA was cloned in the PstI site of pBR322 using the GC tailing procedure. Bacterial transformants were screened by colony hybridization in two steps: first for the presence of sequences complementary to the 280 bp SmaI/BstN1 subfragment of

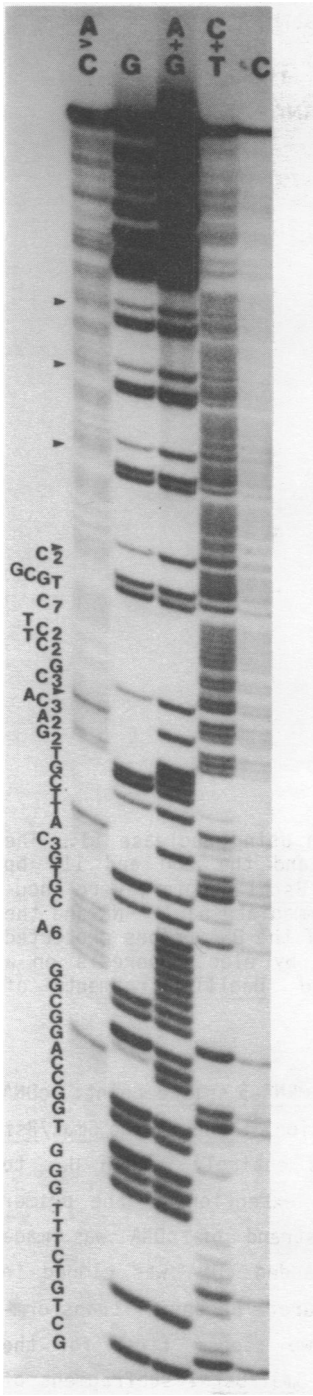


Figure 3. Autoradiograph of a gel used to sequence the 280 bp SmaI/BstNI subfragment. The sequence obtained is shown to to the left of the figure. Arrows represent the end points of the 23bp repeat within the intron.

```
1
GGGGGCCTCC GACGACAGAA ACCCACCGGT CCGCCTTTTT TGCACGGGTA AGCACCTTGG GTGGGCGGAG

                                100
GAGGGGGGGA CGCGGGGGCG GAGGAGGGGG GGACGCGGGG GCGGAGGAGG GGGGGACGCG GGGGCGGAGG

                                                                200
AGGGGGGGAC GCGGGGGCGG AGGAGGGGGC TCACCCGCGT TCGTGCCTTC CCGCAGGAGG AACGCCCTCG

                                                    :
TCGAGGCGAC CGGCGGCGAC CGTTGCGTGG ACCGCTTCCT GCTCGTCGGG:AAAAGCATGT CGTGGGCCCT
TaqI                                               TRS/IRS:US              BstNI
                                                    :
                300
GGAAATGGCG GACACCTTCC TGGACAACAT GCGGGTTGGG CCCAGGACGT ACGCCGACGT ACGCGATGAG
                    BstNI                      BstNI

                                                        400
ATCAATAAAA GGGGGCGTGA GGACCGGGAG GCGGCCAGAA CCGCCGTGCA CGACCCGGAG CGTCCCCTGC

          431
TGCGCTCTCC C
```

Figure 4. Sequence of the non-coding strand of the 431 bp SmaI A fragment. Relevant restriction endonuclease cleavage recognition sequences are underlined, as are the putative ATG initiator codon and the reiterated units within the gene intron. The intron is shown bounded by parentheses, and the junction between $TR_S/IR_S$ and $U_S$ sequences is delineated by a dotted line.

SmaI A and second, for hybridization to the SmaI C fragment that maps adjacent to SmaI A (Figure 1). Hybridization to both probes indicated that the cloned cDNA was extended from the primer across the IE mRNA-5 splice point and into SmaI C fragment sequences, the fragment which encodes the RNA 5' terminus. The two plasmid isolates thus selected (pTH54 and pTH62) were found to contain a single SmaI site, that present between the SmaI A and C fragments, and to contain an insert of approximately 350 bp.

The two cloned cDNA inserts were sequenced from the 5'-$^{32}$P-labeled SmaI site, and the sequences obtained (Figure 5) were compared to that of the SmaI A fragment (Figure 4). The sequences were virtually identical, with the exception that 149 bp, repesenting the intron, were deleted from the cDNAs. This deletion occurred between residues 46-197 (Figure 4). The exact location of the splice point could not be identified as it occured within a series of three G residues. One further difference was noted between

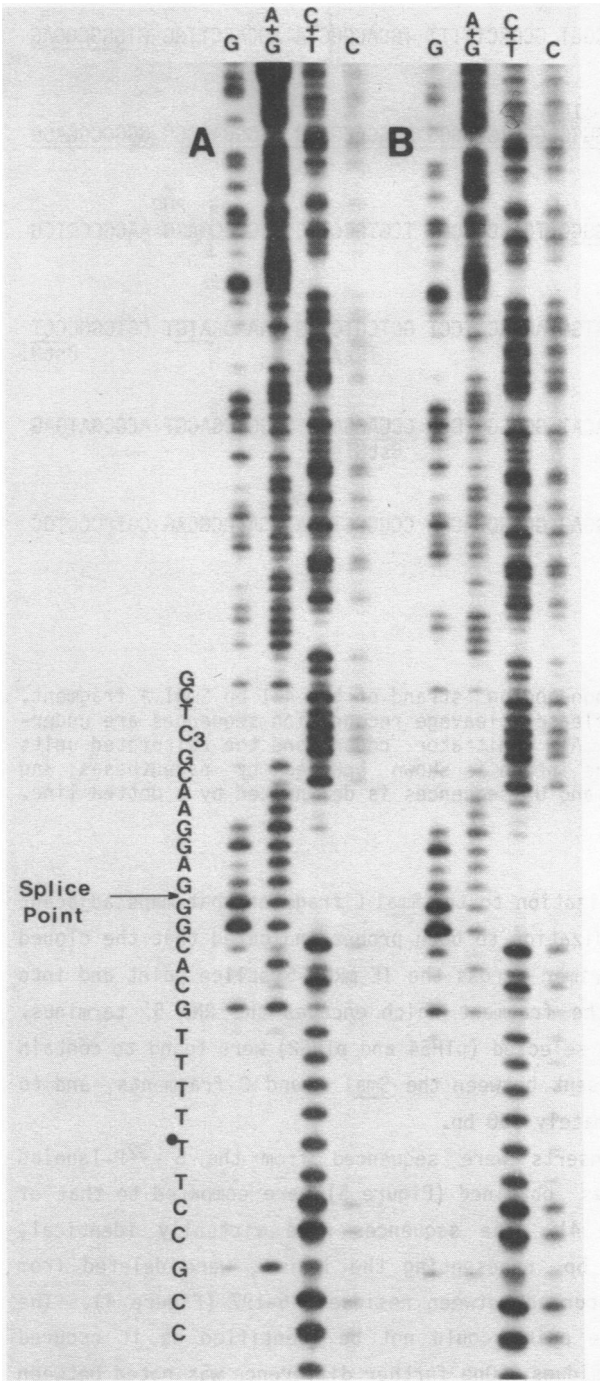


Figure 5. Autoradiogram of a gel used to sequence the IE mRNA-5-specific cDNA clones pTH54 (A) and pTH62 (B). The location of the splice point within the cDNA sequence is indicated by the arrow.

the sequence of pTH62 and that of SmaI A and pTH54: there was a single base substitution at position 213, which destroyed the TaqI site of pTH62.

## DISCUSSION

We have described the sequence of an HSV-1 intron, and have located the RNA splice point within a stretch of three G residues. By using the proposed splicing rules (16,17), the location of the splice point in the cDNA sequence could be deduced. In their simplest form, the rules define the 5' and 3' ends of introns by the dinucleotides GT and AG, respectively (16). This places the 5'exon-intron junction between residues 47-48 and the 3' intron-exon junction between residues 196-197. This splice point is illustrated in the cDNA sequence (Figure 5). The distance from the rightmost BstN1 site (Figure 1) to the 3' end of the intron is, therefore, 81 b; an estimate of 80 b was obtained in the nuclease S1 analysis.

The sequences around the intron-exon junctions are consistent with those described for other eucaryotic genes (17). At the 5' or donor splice site of the IE mRNA-5 gene we found the sequence GGTAAG, which is consistent with two of the four prefered forms: -Pu↓GTAxG and -Pu↓GTxAG, where ↓ denotes the cleavage site (17). At the 3' or acceptor splice site, the sequence CCGCAG is found, consistent with the preferred form PyPyxPyAG↓ (17). Also, in accordance with these rules, 5' to the acceptor splice site is a pyrimidine rich sequence, TTCGTGCCTTCCCGCAG↓, within which the dinucleotide AG is present once only.

The four copies of the 23 bp repeat plus the 15 bp incomplete repeat account for 107 bp of the 149 bp intron. The intron sequences bounding these repeats can largely be recognized as being important for splicing recognition. For example, the 25 bases between the end of the repeating units and the dinucleotide AG are pyrimidine rich (18 of 25), an apparent requirement for splicing (17). Therefore, the IE mRNA-5 intron structure consists primarily of repeating units bounded by sequences required for splicing.

Sequences within the intron, other than those at the intron-exon junction, do not appear to influence splicing (18). It has been suggested, however, that a minimum size (no less than 66 nucleotides) for the intron may be required to allow folding of the RNA so that splicing may occur (19). This minimum size requirement does not explain the reiteration of intron sequences, because nonrepeated intron sequences plus one copy of the 23 bp repeat total 65 nucleotides.

It has been reported (20) that the octanucleotide GGXGGAG is found

within the origins of replication of the papovaviruses BKV, SV40 and polyoma virus (the sequence GGGCGGPuPu is repeated six times within a 62 nucleotide segment of SV40). This sequence is also found within the terminal 100 nucleotides thought to have a role in DNA replication of a number of adenoviruses (21). While it is not clear if this sequence is indicative of a DNA replication origin, and such an octanucleotide would be predicted to occur many times within the 150 kilobase GC-rich HSV-1 DNA, it may be of interest to note that the sequence GGGCGGAG is repeated five times within the intron reiterated units.

It may be expected that the intron reiterated unit structure would be inherently unstable, due perhaps to unequal crossing-over in recombination, both in the virus DNA and when fragments containing this sequence are cloned in plasmid pBR322. Variations in size of the B4R1 fragment and of the SmaI A subfragment have not been noted, however, either in virus plaque isolates or in E. coli (22). Indeed, sequencing an independently cloned SmaI fragment containing the IE mRNA-4 intron has revealed the same number, and identical nucleotide sequence, of the reiterated units within this intron (our unpublished observations).

The sequencing data presented here, and further unpublished data, have revealed no potential initiator codons within the IE mRNA-5 5' leader sequence. The first ATG is found six bases into $U_S$ sequences (Figure 4), from which the reading frame is open for 87 codons. This is consistent with the observation that this mRNA specifies a polypeptide of approximately 12,000 molecular weight (4). Removal of the intron does not, then, affect the reading frame of this mRNA; rather, it places the putative initiator codon closer to the mRNA terminus. It has been reported that splicing is important for stabilization of mRNAs (23,24), possibly as a result of concurrent transport from the nucleus to the cytoplasm. There appears to be no such requirement for the stabilization of the unspliced HSV-1 thymidine kinase mRNA (25), which is synthesized early post-infection,after the onset of virus protein synthesis (26). It remains to be determined, therefore, whether splicing is obligatory for stabilization of immediate-early HSV-1 mRNAs and, if so, what factors influence maturation and stability of unspliced mRNA.

ACKNOWLEDGEMENTS

We thank Dr. George F. Vande Woude for his encouragement and critical reading of this manuscript. K.U. acknowledges the advice and assistance of Dr. K. Mizuuchi during initial DNA sequencing experiments.

REFERENCES


1. Clements, J.B., Watson, R.J. and Wilkie, N.M. (1977) Cell 12, 275-285.
2. Frenkel, N. and Roizman, B. (1972) Proc. Natl. Acad. Sci. U.S.A. 69, 2654-2658.
3. Swanstrom, R.I. and Wagner, E.K. (1974) Virology 60, 522-533.
4. Watson, R.J., Preston, C.M. and Clements, J.B. (1979) J. Virol. 31, 42-52.
5. Clements, J.B., McLauchlan, J. and McGeoch, D.J. (1979) Nucleic Acids Res. 7, 77-91.
6. Watson, R.J., Sullivan, M. and Vande Woude G.F. (1981) J. Virol. 37, 431-444.
7. Maxam, A. and Gilbert, W. (1980) in Methods in Enzymology, Grossman, L. and Moldave, K. Eds, Vol. 65, pp. 499-560, Academic Press, New York.
8. Chalberg, M.D. and Englund, P.T. (1980) in Methods in Enzymology, Gross man, L. and Moldave, K. Eds., Vol. 65, pp. 39-43, Academic Press, New York.
9. Roychoudhury, R. and Wu, R. (1980) in Methods in Enzymology, Grossman, L. and Moldave, K. Eds., Vol. 65, pp. 43-62, Academic Press, New York.
10. Goodman, H.M. and MacDonald, R.J. (1979) in Methods in Enzymology, Wu, R. Ed., Vol 68, pp 75-90, Academic Press, New York.
11. Efstratiadis, A., Kafatos, F.C., Maxam, A.M. and Maniatis, T. (1976) Cell 7, 279-288.
12. Gough, N.M., Webb, E.A., Cory, S. and Adams, J.M. (1980) Biochemistry 19, 2702-2710.
13. Nelson, T. and Brutlag, D. (1979) in Methods in Enzymology, Wu, R. Ed., Vol. 68, pp. 41-50, Academic Press.
14. Denniston, K., Madden, M.J., Enquist, L.W. and Vande Woude, G.F. (1981) Gene, in press.
15. Grunstein, M. and Hogness, D.S. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 3961-3965.
16. Breathnach, R. Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978) Proc. Natl. Acad. Sci. USA. 75, 4853-4857.
17. Seif, I., Khoury, G. and Dhar, R. (1979) Nucleic Acids Res. 6, 3387-3398.
18. Khoury, G., Gruss, P., Dhar, R. and Lai, C.-J. (1979) Cell 18, 85-92.
19. Rogers, J. and Wall, R. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 1877-1879.
20. Seif, I., Khoury, G. and Dhar, R. (1979) Cell 18, 963-977.
21. Tolun, A., Alestrom, P. and Pettersson, U. (1979) Cell 17, 705-713.
22. Umene, K. and Enquist , L.W. (1981) Gene, in press.
23. Hamer, D.H. and Leder, P. (1979) Cell 18, 1299-1302.
24. Gruss, P., Lai, C.-J., Dhar, R. and Khoury, G. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 4317-4321.
25. McKnight, S.L. (1980) Nucleic Acids Res. 8, 5949-5964.
26. Preston, C.M. (1979) J. Virol. 29, 275-284.